

К. О. БАЗІЛЕВИЧ, Є. С. МЕНЯЙЛОВ, Д. І. ЧУМАЧЕНКО

## ВИДІЛЕННЯ ЗОН РОЗПОВСЮДЖЕННЯ ЗАХВОРЮВАНOSTI НА КОРОНАВІРУС COVID-19 НА ОСНОВІ МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ

**Предмет:** використання математичного апарату нейронних мереж для наукового обґрунтування протиепідемічних заходів з метою зниження захворюваності при прийнятті ефективних управлінських рішень. **Ціль:** застосувати кластерний аналіз, на основі нейронної мережі, для вирішення задачі виділення зон розповсюдження захворюваності. **Задачі:** проаналізувати методи аналізу даних для вирішення задачі кластеризації; розробити нейромережовий метод кластеризації територій України за характером епідемічного процесу COVID-19; на основі розробленого методу реалізувати програмний додаток аналізу даних для виділення зон розповсюдження захворюваності на прикладі коронавірусу COVID-19. **Методи:** моделі та методи аналізу даних, моделі та методи теорії систем (на базі інформаційного підходу), методи машинного навчання, зокрема метод Adaptive Boosting (на основі методу градієнтного спуску), методи навчання нейронних мереж. **Результати:** були використані розподілені по областях України дані Центру громадського здоров'я МОЗ України про захворюваність на COVID-19, кількість лабораторно обстежених осіб, кількість проведених лабораторних досліджень методами ПЦР та ІФА, кількість проведених лабораторних досліджень IgA, IgM, IgG; в моделі використані дані з березня 2020 по грудень 2020, при моделюванні не враховані дані з тимчасово окупованих територій України; для кластерного аналізу побудована нейронна мережа з 60 вхідними нейронами, 100 прихованими нейронами з активаційною функцією Фермі та 4 вихідними нейронами; для програмної реалізації моделі використана мова програмування Python. **Висновки:** проведено аналіз методів побудови нейронних мереж; аналіз методів навчання нейронних мереж, у тому числі методу градієнтного спуску; всі теоретичні відомості, описані в цій роботі, були використані для реалізації програмного продукту обробки даних тестування на COVID-19 в Україні; було проведено розбиття областей України на зони зараження вірусом COVID-19 та представлено карту цього розбиття.

**Ключові слова:** кластерний аналіз; нейронна мережа; машинне навчання; епідемічний процес; COVID-19.

### Вступ

Мінімізація наслідків вірусу COVID-19 сьогодні є основним завданням, яким охоплені всі країни світу. Побудовано безліч прогнозів [1-3], розроблено велику кількість моделей поширення захворювання [4-6], епідемічної діагностики, але не дивлячись на спад темпів зараження епідемія вірусу триває.

За даними університету Джона Хопкінса [7], Україна займає 17 сходинку в топі лідерів за виявленими випадками вірусу COVID-19.

Виділення зон розповсюдження захворюваності є важливим завданням у рамках мінімізації наслідків хвороби та зменшення кількості заражень. Проте виділення таких зон лише за територіальним принципом (що зараз використовується в Україні) не є дієвим, адже при такому підході до розгляду не береться реальна ситуація із захворюваності та не аналізуються реальні статистичні дані.

Для вирішення цієї задачі у роботі запропоновано використовувати методи Data Mining, що показують відмінні результати при роботі зі стохастичними процесами [8-10], зокрема кластерний аналіз. Основна ідея кластерного аналізу - виділення серед безлічі даних груп, всередині яких елементи в певній мірі схожі.

Під кластерами у даному випадку маються на увазі зони з однаковим характером епідемічного процесу COVID-19. Таке розбиття дозволить вжити ефективних протиепідемічних заходів зі зниження епідемічної захворюваності на обраних територіях.

В рамках такого аналізу відбувається певна класифікація досліджуваних даних за рахунок

розподілу їх по групах. Ці групи впорядковані ієрархічно і структуру таких одержаних після аналізу кластерів можна представити у вигляді дерева. Для реалізації процедури кластеризації було використано багатопланову нейронну мережу із зворотнім розповсюдженням.

### Аналіз останніх досліджень і публікацій

Застосування кластерного аналізу для вирішення задачі виділення зон розповсюдження захворюваності є важливим науковим завданням, що пропонується вирішувати із застосуванням нейронної мережі.

Під час практичної реалізації нейронної мережі виникає ряд проблем: зазвичай заздалегідь не відомо, якої складності необхідно буде розробити мережу для точної реалізації відображення необхідного процесу [11-13]. Ця складність може виявитися надмірно високою, що буде потребувати складної архітектури такої мережі. Як відомо – одношарові нейронні мережі здатні вирішувати тільки лінійно роздільні завдання [14]. Для зняття цього обмеження у роботі пропонується використовувати багатопланові нейронні мережі [15,16]. В мережі з одним прихованим шаром вектор, відповідний до вхідного зразку, перетворюється прихованим шаром в певний новий простір, що може мати іншу розмірність, а потім гіперплощини, відповідні нейронам вихідного шару, поділяють його на класи. Таким чином, мережа розпізнає не тільки характеристики вихідних даних, але і "характеристики характеристик", сформовані прихованим шаром.

У багатопланових перцептронах кожен нейрон має нелінійну функцію активації. Ця нелінійна функція є гладкою. Найпопулярнішою формою

функції, що задовольняє цій вимозі, є сигмоїдальна (що і використовується у дослідженні). Наявність нелінійності грає дуже важливу роль, так як в іншому випадку відображення "вхід-вихід" мережі можна звести до звичайного одношарового персептрону. Більш того, використання логістичної функції мотивовано біологічно, так як в ній враховується відновлювальна фаза реального нейрона. Мережа містить один або кілька шарів прихованих нейронів, які не є частиною входу або виходу мережі. Ці нейрони дозволяють мережі навчатися вирішенню складних завдань, послідовно отримуючи найбільш важливі ознаки з вхідного образу (вектора). Мережа має високий ступінь зв'язності, що реалізовується за допомогою синаптичних з'єднань [17]. Саме цей вид нейронної мережі було обрано для досягнення мети дослідження.

**Метою статті** є розробка програмного додатку аналізу даних для виділення зон розповсюдження захворюваності на основі методів кластерного аналізу (на прикладі захворюваності на коронавірус COVID-19).

Дана постановка відповідає логіці побудови нейронної мережі для кластеризації даних, що дозволяє досягти потрібного результату. Результатом вирішення даного завдання буде розподіл України на кластери за станом захворюваності (що відповідає зонам карантину).

### Основна частина

Розглянемо постановку задачі кластеризації. Нехай  $X$  – множина об'єктів,  $Y$  – множина номерів кластерів. Задана функція відстані між об'єктами  $\rho(x, x')$ . Є кінцева навчальна вибірка об'єктів

$X^m = \{x_1, x_2, \dots, x_m\} \subset X$ . Потрібно розбити вибірку на непересічні підмножини, які називаються кластерами, так, щоб кожен кластер складався з об'єктів, близьких за метрикою  $\rho$ , а об'єкти різних кластерів істотно відрізняються. При цьому кожному об'єкту  $x_i \in X^m$  задається номер кластера  $y_i$ . Алгоритм кластеризації – це функція  $a: X \rightarrow Y$ , що будь-якому об'єкту  $x \in X$  ставить у відповідність номер кластера  $y \in Y$ . Множина  $Y$  в деяких випадках відома заздалегідь, однак частіше ставиться завдання визначити оптимальне число кластерів, з точки зору того чи іншого критерію якості кластеризації. Кластеризація (навчання без вчителя) відрізняється від класифікації (навчання з учителем) тим, що мітки вихідних об'єктів  $y_i$  спочатку не задані, і навіть може бути невідома сама множина  $Y$ .

Кластеризація за допомогою нейронної мережі виконується наступним чином. На вхід до нейронної мережі подається вхідний вектор ознак: відповідно до алгоритму зворотного розподілу помилки, відбувається коригування ваг згідно заздалегідь

зумовленої класу вхідного вектора. Після навчання на виділення кількості вхідних образів відбувається перевірка мережі на певній кількості вхідних образів, які не брали участі в навчанні.

### Метод виділення зон розповсюдження захворюваності на основі процедури градієнтного спуску

Процедури градієнтного спуску зазвичай використовуються там, де потрібно максимізувати або мінімізувати  $n$ -вимірні функції.

Градiєнт – це вектор  $g$ , який визначений для будь-якої диференційованої точки функції, який вказує від цієї точки точно до найкрутішого підйому і вказує градієнт у цьому напрямку за допомогою своєї норми  $|g|$ . Таким чином, градієнт є узагальненням похідної для багатовимірних функцій. Відповідно, негативний градієнт  $-g$  точно вказує на найкрутіший спуск. Оператор градієнта  $\nabla$  називається оператором набла, загальний запис градієнта  $g$  точки  $(x, y)$  двовимірної функції  $f$  становить  $g(x, y) = \nabla f(x, y)$ .

Нехай  $g$  – градієнт. Тоді  $g$  – вектор із  $n$  компонентами, який визначений для будь-якої точки (диференціальної)  $n$ -мірної функції  $f(x_1, x_2, \dots, x_n)$ .

Позначення оператора градієнта визначається як  $g(x_1, x_2, \dots, x_n) = \nabla f(x_1, x_2, \dots, x_n)$ .  $g$  спрямовує з будь-якої точки в бік найкрутішого підйому з цієї точки, з  $|g|$  що відповідає ступеню цього підйому. Нехай  $f$  –  $n$ -мірна функція, а  $s = (s_1, s_2, \dots, s_n)$  задана початкова точка. Градієнтний спуск означає перехід від  $f(s)$  проти напрямку  $g$ , тобто до  $-g$  із кроками розміром  $|g|$  у напрямку до все менших і менших значень  $f$ .

Зворотне розповсюдження – це процедура градієнтного спуску з функцією помилки  $Err(W)$ , яка отримує всі  $n$  вагових коефіцієнтів як аргументи та присвоює їм вихідні помилки, тобто стає  $n$ -мірною. На  $Err(W)$  за допомогою градієнтного спуску шукається точка невеликої помилки або навіть точка найменшої помилки.

Таким чином, аналогічно правилу дельта, зворотне розмноження тренує ваги нейронної мережі. І саме правило дельти або його змінна  $\delta_i$  для нейрона і розширюється від одного тренованого вагового шару до декількох шляхом зворотного розповсюдження.

Вхід мережі в окремі нейрони і є результатом зваженої суми. Крім того, як і при виведенні правила дельта, нехай  $o_{p,i}$ ,  $net_{p,i}$  т. д. Визначаються як вже знайомі  $o_i$ ,  $net_i$  тощо за шаблоном введення  $p$ , який використаний для навчання.

Нехай вихідна функція знову буде тотожністю, отже,  $o_i = f_{act}(net_{p,i})$  виконується для будь-якого нейрона  $i$ .

Тепер виконаємо те саме виведення, що і для правила дельта та функцій розбиття за допомогою

правила ланцюга. Спочатку потрібно вивести функцію помилки  $Err$  відповідно до ваги  $w_{k,h}$ .

$$\frac{\partial Err(w_{k,h})}{\partial w_{k,h}} = \frac{\partial Err}{\partial net_h} \cdot \frac{\partial net_h}{\partial w_{k,h}} \cdot \frac{\partial Err}{\partial net_h} = -\delta_h.$$

Першим фактором рівняння є  $-\delta_h$ , який буде розглянуто пізніше. Чисельник другого множника рівняння включає вхід мережі, тобто зважена сума включається в чисельник, щоб можна було негайно його отримати. Знову усі доданки суми випадають, крім суми, що містить  $w_{k,h}$ . Цей доданок називається  $w_{k,h} \cdot o_k$ . Якщо обчислити похідну, вихід нейрона  $k$  стає:

$$\frac{\partial net_h}{\partial w_{k,h}} = \frac{\partial \sum_{k \in K} w_{k,h} o_k}{\partial w_{k,h}} = o_k,$$

$$\delta_h = -\frac{\partial Err}{\partial net_h} = -\frac{\partial Err}{\partial o_h} \cdot \frac{\partial o_h}{\partial net_h}.$$

Згідно з визначенням багатовимірного ланцюгового правила, одразу отримаємо рівняння

$$-\frac{\partial Err}{\partial o_h} = \sum_{l \in L} \left( -\frac{\partial Err}{\partial net_l} \cdot \frac{\partial net_l}{\partial o_h} \right).$$

Сума у верхньому рівнянні містить два фактори. Обчислимо другий коефіцієнт у наступному рівнянні

$$\frac{\partial net_l}{\partial o_h} = \frac{\partial \sum_{h \in H} w_{h,l} \cdot o_h}{\partial o_h} = w_{h,l}.$$

Те саме стосується першого фактора згідно з визначенням  $\delta$ :

$$-\frac{\partial Err}{\partial net_l} = \delta_l.$$

Тепер:

$$-\frac{\partial Err}{\partial o_h} = \sum_{l \in L} \delta_l w_{h,l}$$

тому бажана зміна ваги  $\Delta w_{k,h}$

$$\Delta w_{k,h} = \eta o_k \delta_h,$$

$$\delta_h = f'_{act}(net_h) \cdot \sum_{l \in L} (\delta_l w_{h,l})$$

звичайно лише у випадку, коли  $h$  є внутрішнім нейроном (інакше не було б подальшого шару  $L$ ). Загалом, результатом є узагальнення правила дельта, яке називається зворотним розповсюдженням помилки:

$$\Delta w_{k,h} = \eta o_k \delta_h,$$

$$\delta_h = \left\{ \begin{array}{l} f'_{act}(net_h) \cdot (t_h - y_h) \\ f'_{act}(net_h) \cdot \sum_{l \in L} (\delta_l w_{h,l}) \end{array} \right\}.$$

На відміну від правила дельти,  $\delta$  трактується по-різному, залежно від того, чи  $h$  є вихідним чи внутрішнім (тобто прихованим) нейроном:

Якщо  $h$  – вихідний нейрон, то

$$\delta_{p,h} = f'_{act}(net_{p,h}) \cdot (t_{p,h} - y_{p,h}).$$

Таким чином, за означенною схемою навчання  $p$  вага  $w_{k,h}$  від  $k$  до  $h$  пропорційно змінюється відповідно до

- швидкості навчання  $\eta$ ;
- вихідного  $o_{p,k}$  нейрона-попередника  $k$ ;
- градієнта функції активації в положенні мережевого входу нейрона-наступника  $f'_{act}(net_{p,h})$ ;
- різниці між вводом  $t_{p,h}$  та виходом  $y_{p,h}$  нейрона-наступника  $h$ .

У цьому випадку зворотне розмноження працює на двох нейронних шарах, вихідному шарі з наступним нейроном  $h$  і попереднім шаром з попередником нейроном  $k$ . Якщо  $h$  – внутрішній, прихований нейрон, то має місце.

$$\delta_{p,h} = f'_{act}(net_{p,h}) \cdot \sum_{l \in L} (\delta_{p,l} \cdot w_{h,l}).$$

Опис результатів, отриманих при роботі з програмним додатком

В якості вхідних даних є дата-сет зі статистикою проходження пацієнтів тестування на Covid-19 угрупований по областям України та медичним закладам [18]. Повний список параметрів:

- рівень зараженості;
- область;
- назва закладу (установи);
- кількість лабораторно обстежених осіб (первинні та ретестування);
- кількість проведених лабораторних досліджень методом ПЛР на COVID-19;
- кількість проведених лабораторних досліджень хворих з пневмоніями;
- кількість позитивних результатів на COVID-19;
- кількість результатів тестувань на COVID-19;
- кількість зразків, що надійшли на дослідження ПЛР;
- кількість непротестованих зразків, що залишаються в лабораторії;
- кількість лабораторно обстежених осіб;
- кількість проведених лабораторних досліджень;
- кількість позитивних результатів досліджень;
- виявлення антигенів.

На рис. 1 можна побачити частину імпортованих з дата-сету даних.

Результат нормалізації імпортованих даних можна побачити на рис. 2. Як можна бачити перші два стовпця відсутні через те що вони мають строковий тип даних. Усі 60 значень кожного запису представляють вхідний нейрон мережі.

Розглянемо приклад роботи програми з моделлю нейронної мережі, вказаною на рис. 3.

	2	3	4	...	58	59	60
ДУ «Вінницький ОЛЦ МОЗ України»	1316	93	...	209	8	0	
ТОВ "Діагностичний центр "МЕДІЛАБС", м. Вінниця	350	0	...	1986	0	0	
КНП "Вінницький обласний клінічний центр профі...	94	4	...	2991	87	47	
КНП "Вінницька обласна клінічна лікарня ім.М.І...	0	0	...	591	188	0	
КНП "Вінницький обласний клінічний високоспеці...	0	0	...	298	26	0	
...	...	...	...	...	...	...	
КНП "ФТИЗИАТРІЯ"	0	0	...	15	15	0	
КНП "КМДКЛ № 2"	0	0	...	36	26	0	
ТОВ Нова діагностика	0	0	...	140	0	0	
КНП "Київський міський пологовий будинок №6"	0	0	...	0	0	0	
ТОВ "Клініка Маркова"	22	0	...	0	0	0	

[umns]

Рис. 1. Імпортовані данні

	3	4	5	6	7
0	0.299295	0.162872	0.192771	0.302479	0.203390
1	0.079600	0.000000	0.000000	0.082101	0.000000
2	0.021378	0.007005	0.000000	0.021833	0.014124
3	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000
...	...	...	...	...	...
455	0.000000	0.000000	0.000000	0.000000	0.000000
456	0.000000	0.000000	0.000000	0.000000	0.000000
457	0.000000	0.000000	0.000000	0.000000	0.000000
458	0.000000	0.000000	0.000000	0.000000	0.000000
459	0.005003	0.000000	0.000000	0.005003	0.000000

[2300 rows x 60 columns]

Рис.2. Нормалізовані данні

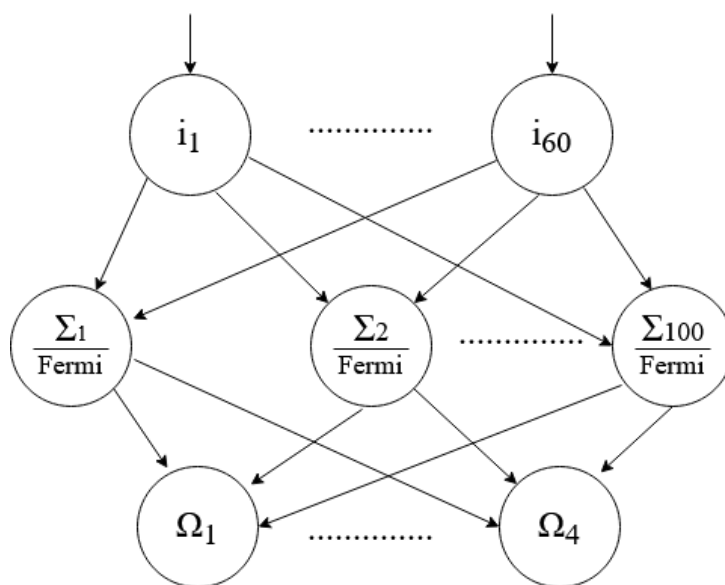


Рис. 3. Модель нейронної мережі з 60 входними нейронами, 100 прихованими нейронами з функцією активації Фермі та 4 вихідними нейронами

Вихідний нейрон – вектор з чотирьох компонент, індекс максимального значення якого є індексом кластеру до якого належить запис.

Між входними прихованими нейронами міститься матриця вагових коефіцієнтів, яка на першій ітерації ініціалізується випадковими значеннями. Розмір цієї

матриці (60x100). Вагові коефіцієнти першого входного нейрону до усіх нейронів на прихованому рівні можна побачити на рис. 4. Кожне значення матриці коефіцієнтів буде змінюватись у період навчання.

```

[-0.55074841 -0.30045617 0.40008079 -0.62892543 -0.552295
 1.79613424 0.20481614 0.3304295 0.34907633 2.188642
 0.78918423 -2.04289116 -0.04020203 1.40965998 -0.356157
 2.17095693 0.09459189 -0.18997519 0.40140887 -0.989635
 0.60683112 -0.07350352 0.07641064 0.80437461 0.629701
 1.10081371 0.29816912 -1.18169771 -0.09301111 0.628639
 -0.15680307 0.12327033 0.41433496 -0.28864685 2.244296
 -0.0559052 -0.26257586 0.68607464 1.86950358 1.208992
 -1.9326388 -0.86330387 -0.64293224 0.09058302 0.405786
 0.16269968 -0.88488518 -0.61878525 -1.36440879 -0.888033
 0.18901213 0.63968365 -0.97510193 0.40153046 -0.511753
 -0.86328308 -1.91894846 1.17076114 -0.19605497 0.682347
 -1.61067836 -1.5535064 -0.99265388 -1.52559155 2.404002
 -0.54133163 0.06468708 -1.41063762 -1.02973355 -2.159183
 0.36913413 0.12705528 -1.88545164 1.19693911 -0.177951
 0.44241701 0.82330423 0.45288699 -0.27477905 -0.585110
 0.889565 -0.65124859 -2.29598436 0.60954192]

```

Рис. 4. Вагові коефіцієнти першого вхідного нейрону до усіх нейронів на прихованому рівні

На рис. 5 представлено графік значення функції похибки на кожній ітерації навчання. З графіку помітно, що приблизно з 350 епохи навчання,

значення функції похибки змінюється незначним чином, тому подальше навчання нейронної мережі не має сенсу.

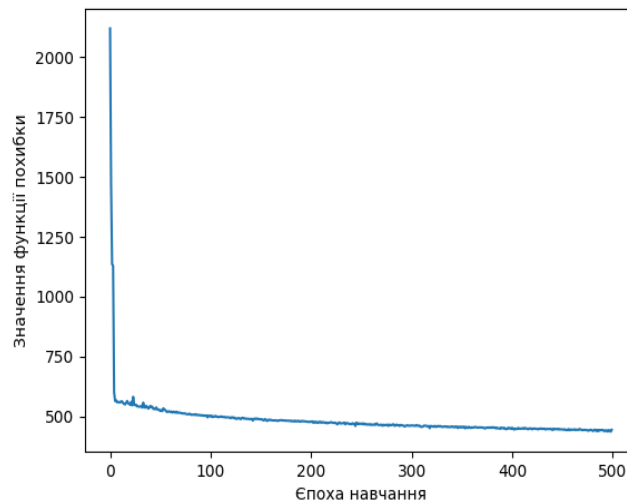


Рис. 5. Значення функції похибки на кожній ітерації навчання

На рис. 6 можна побачити графік відношення кількості вірних прогнозів до загальної кількості даних (точність прогнозу) на кожній ітерації навчання.

Як можна бачити на рис. 6 чим більше

епох навчання, тим вище точність прогнозу нейронної мережі.

На рис. 7 представлено розподілення областей України на області зараження вірусом Covid-19 мережею.

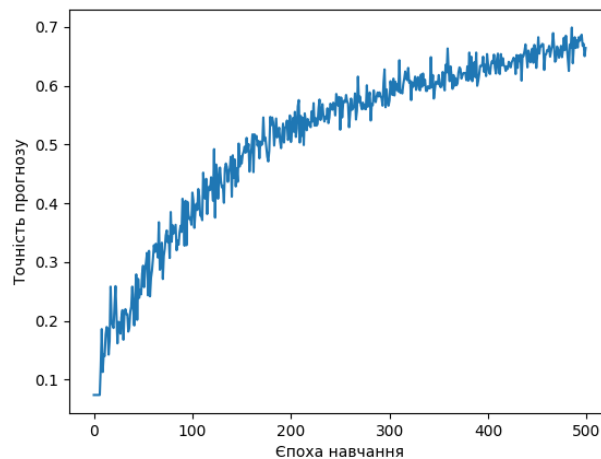


Рис. 6. Графік точності прогнозу на кожній ітерації навчання



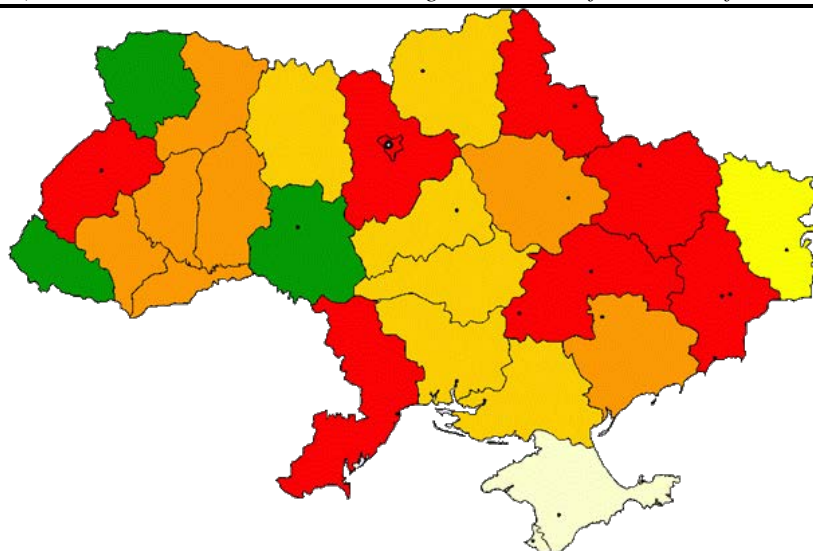


Рис. 7. Розбиття областей України на зони зараження

На рис. 7 області України замальовані кольорами відносно до їх перебування у зонах зараження вірусом COVID-19 (зелена, жовта, помаранчева, червона). Ціль моделювання була досягнута. На рис. 8 можна бачити приклад вихідного шару нейронної мережі.

Результати кластерного аналізу показали, що динаміку розповсюдження захворюваності на COVID-19 в Україні можна розділити на 4 кластери за схожістю епідемічного процесу. До першого кластеру

увійшли Волинська, Закарпатська та Вінницька області. До другого кластеру увійшли Львівська, Одеська, Сумська, Харківська, Донецька, Дніпропетровська, Київська області та місто Київ. Третій кластер включив Рівненську, Хмельницьку, Тернопільську, Чернівецьку, Івано-Франківську, Полтавську та Запорізьку області. Четвертий кластер включив всі інші області України.

```
[[[9.97716555e-01 9.98472477e-01 9.99236074e-01 5.63135028e-01]
[3.76436408e-01 9.99999999e-01 1.00000000e+00 9.99394698e-01]
[9.96782750e-01 9.98576014e-01 9.99305480e-01 5.23992579e-01]
[7.35402156e-01 9.99253078e-01 9.99923176e-01 4.53543932e-04]
[9.90967042e-01 9.98862561e-01 9.99263636e-01 6.99411258e-01]
[9.98845095e-01 9.97543204e-01 9.98503758e-01 4.25000149e-01]
[9.99957668e-01 9.97566808e-01 9.97340075e-01 6.88395617e-02]
[9.98258399e-01 9.99499650e-01 9.98970073e-01 2.61586687e-01]
[9.96889145e-01 9.98350235e-01 9.99298159e-01 5.33690463e-01]
[9.89137103e-01 9.98523057e-01 9.99627291e-01 5.30845169e-01]
[9.97380185e-01 9.97937991e-01 9.99061508e-01 6.14148761e-01]
[9.97114161e-01 9.98896633e-01 9.99384229e-01 5.19558518e-01]
[9.98387103e-01 9.98346917e-01 9.98770854e-01 4.45148473e-01]
[9.98105187e-01 9.98147540e-01 9.99183996e-01 5.54695172e-01]
[9.96732737e-01 9.99167121e-01 9.99427817e-01 5.90957768e-01]
[6.43734464e-06 9.99956512e-01 9.99993872e-01 9.99999990e-01]
[9.97445042e-01 9.96305550e-01 9.98784347e-01 7.83310043e-01]
[9.97657331e-01 9.96024236e-01 9.98655196e-01 7.95962663e-01]
[9.97116536e-01 9.96753021e-01 9.99208027e-01 7.97290689e-01]
[6.85970365e-01 9.98541120e-01 9.99993736e-01 9.95937710e-01]
[9.97971758e-01 9.97350963e-01 9.96068559e-01 1.29123515e-01]
[9.96210194e-01 9.99525166e-01 9.99190385e-01 4.07816005e-01]
[9.99507134e-01 9.96863581e-01 9.99385814e-01 4.43537046e-03]
[9.99972151e-01 9.98021060e-01 9.95538686e-01 7.99327279e-01]
[9.97220304e-01 9.98454135e-01 9.99033283e-01 3.71124277e-01]
[9.93911573e-01 9.99979428e-01 9.99999961e-01 9.99998772e-01]
[9.99877240e-01 9.99994058e-01 9.99421883e-01 5.52667613e-02]
[9.98032586e-01 9.98421984e-01 9.99193511e-01 4.92604748e-01]
[9.97021522e-01 9.98903128e-01 9.99546831e-01 4.76057802e-01]
[9.99298878e-01 9.99143813e-01 9.98291324e-01 3.71522259e-01]
[9.99141860e-01 9.98290020e-01 9.91545705e-01 1.20325047e-01]
[9.93698151e-01 9.98810087e-01 9.99673666e-01 6.63315841e-01]]]
```

Рис. 8. Приклад вихідного шару нейронної мережі

Результатом є максимальне число з вектору чотирьох компонент. Такий результат є унікальним

тому що враховує реальні дані захворюваності на Covid-19.

### Висновки

В роботі було розглянуто та вирішено такі завдання: аналіз методів побудови нейронних мереж; аналіз методів навчання нейронних мереж, у тому числі метод градієнтного спуску; розробка програмного додатку аналізу даних для виділення зон розповсюдження захворюваності на основі методів кластерного аналізу (на прикладі захворюваності на коронавірус COVID-19).

Всі теоретичні відомості, описані в цій роботі, були використані для реалізації програмного продукту

обробки даних тестування на COVID-19 в Україні. Було проведено розбиття областей України на зони зараження коронавірусом COVID-19 та представлено графік цього розбиття.

\*Дослідження виконане в рамках проєкту Національного фонду досліджень України №2020.02/0404 "Розробка інтелектуальних технологій оцінки епідемічної ситуації для підтримки прийняття управлінських рішень у сфері біобезпеки населення" [19].

### Список літератури

1. Tabik S., et al. COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images. *IEEE Journal of Biomedical and Health Informatics*. 2020. Vol. 24. No. 12. P. 3595–3605. DOI: <https://doi.org/10.1109/JBHI.2020.3037127>
2. Marmarelis V. Z. Predictive Modeling of Covid-19 Data in the US: Adaptive Phase-Space Approach. *IEEE Open Journal of Engineering in Medicine and Biology*. 2020. Vol. 1. P. 207–213. DOI: <https://doi.org/10.1109/OJEMB.2020.3008313>
3. Cihan P. Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak. 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). Istanbul, Turkey, 2020. P. 1–4. DOI: <https://doi.org/10.1109/ISMSIT50672.2020.9254714>.
4. Barman M., Mishra N. A time-delay SEAIR model for COVID-19 spread. IEEE 4th Conference on Information & Communication Technology (CICT). Chennai, India, 2020. P. 1–6. DOI: <https://doi.org/10.1109/CICT51604.2020.9312111>
5. Kapetanović A. L., Poljak D. Modeling the Epidemic Outbreak and Dynamics of COVID-19 in Croatia. 5th International Conference on Smart and Sustainable Technologies (SpliTech). Split, Croatia, 2020. P. 1–5. DOI: <https://doi.org/10.23919/SpliTech49282.2020.9243757>
6. Horry M. J., et al. COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data. *IEEE Access*. 2020. Vol. 8. P. 149808–149824. DOI: <https://doi.org/10.1109/ACCESS.2020.3016780>
7. Johns Hopkins University & Medicine. Coronavirus resource center Baltimore, USA : site. URL: <https://coronavirus.jhu.edu/map.html> (last accessed: 17.02.2021).
8. Bazilevych K., et al. Stochastic modelling of cash flow for personal insurance fund using the cloud data storage. *International Journal of Computing*. 2018. № 3 (17). P. 153–162. DOI: <https://doi.org/10.47839/ijc.17.3.1035>
9. Базилевич К. А., и др. Определение вероятности заболевания болезнями сердца на основе методов Data Mining. *Интегрированные технологии в проектировании и конструировании*. 2019. № 83. С. 202–214.
10. Bazilevych K., et al. Determining the Probability of Heart Disease using Data Mining Methods. *CEUR Workshop Proceedings*. 2019. Vol. 2488. P. 1–12.
11. Fang W., Lacher R. C. Network complexity and learning efficiency of constructive learning algorithms. *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, Orlando, FL, USA, 1994, P. 366–369 Vol. 1. DOI: <https://doi.org/10.1109/ICNN.1994.374191>
12. Tuma A., Haasis H., Rentz O. Emission oriented production control strategies based on fuzzy expert systems, neural networks and neuro-fuzzy approaches. *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, Nagoya, Japan, 1993, pp. 2971–2974 vol.3. DOI: <https://doi.org/10.1109/IJCNN.1993.714346>
13. Bu Z., et al. Encrypted Network Traffic Classification Using Deep and Parallel Network-in-Network Models. *IEEE Access*. 2020. vol. 8, pp. 132950–132959. DOI: <https://doi.org/10.1109/ACCESS.2020.3010637>
14. Goudreau M. W., et al. First-order versus second-order single-layer recurrent neural networks. *IEEE Transactions on Neural Networks*. May 1994. Vol. 5. No. 3. P. 511–513. DOI: <https://doi.org/10.1109/72.286928>
15. Guha D. R., Patra S. K. Cochannel Interference Minimization Using Wilcoxon Multilayer Perceptron Neural Network. 2010 International Conference on Recent Trends in Information, Telecommunication and Computing. Kerala, India. 2010. P. 145–149. DOI: <https://doi.org/10.1109/ITC.2010.50>
16. Hirahara M., Oka N. A hybrid model composed of a multilayer perceptron and a radial basis function network. *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*. Nagoya, Japan, 1993, P. 1353–1356. Vol. 2. DOI: <https://doi.org/10.1109/IJCNN.1993.716794>
17. Постарнак Д. В. Критический анализ моделей нейронных сетей. *Вестник Тюменского государственного университета*. 2012. № 4. С. 162–167.
18. Аналітичні панелі та відкриті дані: Офіційний канал Міністерства Охорони Здоров'я України : сайт. URL: <https://covid19.gov.ua/analitichni-paneli-dashbord> (дата звернення : 17.02.2021).
19. Yakovlev S., et al. The Concept of Developing a Decision Support System for the Epidemic Morbidity Control. *CEUR Workshop Proceedings*. 2020. Vol. 2753. P. 265–274.

### References

1. Tabik, S., Gomez-Rios, A., Martin-Rodriguez, J., Sevillano-Garcia, I., Rey-Area, M., Charte, D., Guirado, E., Suarez, J., Luengo, J., Valero-Gonzalez, M., Garcia-Villanova, P., Olmedo-Sanchez, E. and Herrera, F. (2020), "COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images", *IEEE Journal of Biomedical and Health Informatics*, No. 24 (12), P. 3595–3605. DOI: <https://doi.org/10.1109/JBHI.2020.3037127>
2. Marmarelis, V. (2020), "Predictive Modeling of Covid-19 Data in the US: Adaptive Phase-Space Approach", *IEEE Open Journal of Engineering in Medicine and Biology*, P. 207–213. DOI: <https://doi.org/10.1109/OJEMB.2020.3008313>

3. Cihan, P. (2020), "Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak", P. 1–4. DOI: <https://doi.org/10.1109/ISMSIT50672.2020.9254714>
4. Barman, M. and Mishra, N. (2020), "A time-delay SEAIR model for COVID-19 spread", P. 1–6. DOI: <https://doi.org/10.1109/CICT51604.2020.9312111>
5. Kapetanović, A. L. and Poljak, D. (2020), "Modeling the Epidemic Outbreak and Dynamics of COVID-19 in Croatia", *5th International Conference on Smart and Sustainable Technologies (SpliTech)*, Split, Croatia, P. 1–5, DOI: <https://doi.org/10.23919/SpliTech49282.2020.9243757>
6. Horry, M., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., Shukla, N. (2020), "COVID-19 Detection through Transfer Learning using Multimodal Imaging Data", *IEEE Access*, Vol. 8, P. 149808–149824. DOI: <https://doi.org/10.1109/ACCESS.2020.3016780>
7. Johns Hopkins University & Medicine, "Coronavirus resource center Baltimore", USA: available at: <https://coronavirus.jhu.edu/map.html> (last accessed 17.02.2021).
8. Bazilevych, K., Mazorchuk, M., Parfeniuk, Y., Dobriak, V., Meniailov, I., Chumachenko, D. (2018), "Stochastic modelling of cash flow for personal insurance fund using the cloud data storage", *International Journal of Computing*, Vol. 17, P. 153–162. DOI: <https://doi.org/10.47839/ijc.17.3.1035>
9. Bazilevych, K., Meniailov, I., Goranina, S., Fedulov, K. (2019), "Determination of the likelihood of heart disease based on Data Mining methods" ["Opredeleniye veroyatnosti zabolevaniya boleznyamy serdca na osnove metodov Data Mining"], *Integrated technologies in design and construction*, Vol. 83, P. 202–214.
10. Bazilevych, K., Meniailov, I., Fedulov, K., Goranina, S., Chumachenko, D., Pyrokhov P. (2019), "Determining the Probability of Heart Disease using Data Mining Methods", *CEUR Workshop Proceedings*, Vol. 2488, P. 1–12.
11. Fang, W. and Lacher, R. C. (1994), "Network complexity and learning efficiency of constructive learning algorithms," *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, Orlando, FL, USA, Vol. 1, P. 366–369. DOI: <https://doi.org/10.1109/ICNN.1994.374191>
12. Tuma, A., Haasis, H. and Rentz, O. (1993), "Emission oriented production control strategies based on fuzzy expert systems, neural networks and neuro-fuzzy approaches", *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, Nagoya, Japan, Vol. 3, P. 2971–2974. DOI: <https://doi.org/10.1109/IJCNN.1993.714346>
13. Bu, Z., Zhou, B., Cheng, P., Zhang, K., Ling, Z. -H. (2020), "Encrypted Network Traffic Classification Using Deep and Parallel Network-in-Network Models", *IEEE Access*, Vol. 8, P. 132950–132959. DOI: <https://doi.org/10.1109/ACCESS.2020.3010637>
14. Goudreau, M. W., Giles, C. L., Chakradhar, S. T., Chen, D. (1994), "First-order versus second-order single-layer recurrent neural networks", *IEEE Trans Neural Netw*, No. 5(3), P. 511–3. DOI: <https://doi.org/10.1109/72.286928>
15. Guha, D. R. and Patra, S. K. (2010), "Cochannel Interference Minimization Using Wilcoxon Multilayer Perceptron Neural Network", *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*, Kerala, India, P. 145–149. DOI: <https://doi.org/10.1109/ITC.2010.50>
16. Hirahara, M. and Oka, N. (1993), "A hybrid model composed of a multilayer perceptron and a radial basis function network", *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, Nagoya, Japan, P. 1353–1356, Vol. 2. DOI: <https://doi.org/10.1109/IJCNN.1993.716794>
17. Postarnak, D. V. (2012), "Critical analysis of neural network models" ["Krytycheskyj analiz modelej nejronnykh setej"], *Bulletin of the Tyumen State University*, No. 4, P. 162–167.
18. "Analytical panels and open data: Official channel of the Ministry of Health of Ukraine", available at: <https://covid19.gov.ua/analitichni-paneli-dashbord> (last accessed 17.02.2021).
19. Yakovlev, S., Bazilevych, K., Chumachenko, D., Chumachenko, T., Huliantskyi, L., Meniailov, I., Tkachenko, A. (2020), "The Concept of Developing a Decision Support System for the Epidemic Morbidity Control", *CEUR Workshop Proceedings*, Vol. 2753, P. 265–274.

Надійшла (Received) 22.02.2021

Відомості про авторів / Сведения об авторах / About the Authors

**Базілевич Ксенія Олексіївна** – кандидат технічних наук, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри математичного моделювання та штучного інтелекту, Харків, Україна, email: k.bazilevych@khai.edu; ORCID: <https://orcid.org/0000-0001-5332-9545>.

**Базілевич Ксения Алексеевна** – кандидат технических наук, Национальный аэрокосмический университет им. Н. Е. Жуковского "Харьковский авиационный институт", доцент кафедры математического моделирования и искусственного интеллекта, Харьков, Украина, email: k.bazilevych@khai.edu; ORCID: <https://orcid.org/0000-0001-5332-9545>.

**Bazilevych Kseniia** – PhD (Information Technologies), National Aerospace University "Kharkiv Aviation Institute", Associate Professor at the Department of Mathematical Modeling and Artificial Intelligence, Kharkiv, Ukraine.

**Меняйлов Євген Сергійович** – Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", старший викладач кафедри математичного моделювання та штучного інтелекту, Харків, Україна, email: j.menyailov@khai.edu; ORCID: <https://orcid.org/0000-0002-9440-8378>.

**Меняйлов Евгений Сергеевич** – Национальный аэрокосмический университет им. Н. Е. Жуковского "Харьковский авиационный институт", старший преподаватель кафедры математического моделирования и искусственного интеллекта, Харьков, Украина, email: j.menyailov@khai.edu; ORCID: <https://orcid.org/0000-0002-9440-8378>.

**Meniailov Ievgen** – National Aerospace University "Kharkiv Aviation Institute", Senior Lecturer at the Department of Mathematical Modeling and Artificial Intelligence, Kharkiv, Ukraine

**Чумаченко Дмитро Ігорович** – кандидат технічних наук, доцент, Національний аерокосмічний університет ім. М. Є. Жуковського "Харківський авіаційний інститут", доцент кафедри математичного моделювання та штучного інтелекту, Харьков, Украина, email: d.chumachenko@khai.edu; ORCID: <https://orcid.org/0000-0003-2623-3294>.



**Чумаченко Дмитрій Игоревич** – кандидат технических наук, доцент, Национальный аэрокосмический университет им. Н. Е. Жуковского "Харьковский авиационный институт", доцент кафедры математического моделирования и искусственного интеллекта, Харьков, Украина.

**Chumachenko Dmytro** – PhD (Systems and Means of Artificial Intelligence), Associate Professor, National Aerospace University "Kharkiv Aviation Institute", Associate Professor at the Department of Mathematical Modeling and Artificial Intelligence, Kharkiv, Ukraine.

## ОПРЕДЕЛЕНИЕ ЗОН РАСПРОСТРАНЕНИЯ ЗАБОЛЕВАЕМОСТЬЮ КОРОНАВИРУСОМ COVID-19 НА ОСНОВЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА

**Предмет:** использование математического аппарата нейронных сетей для научного обоснования противоэпидемических мер с целью снижения заболеваемости при принятии эффективных управленческих решений. **Цель:** применить кластерный анализ, на основе нейронной сети, для решения задачи выделения зон распространения заболеваемости. **Задачи:** проанализировать методы анализа данных для решения задачи кластеризации; разработать нейросетевой метод кластеризации территории Украины по характеру эпидемического процесса COVID-19; на основе разработанного метода реализовать программный продукт анализа данных для выделения зон распространения заболеваемости на примере коронавируса COVID-19. **Методы:** модели и методы анализа данных, модели и методы теории систем (на базе информационного подхода), методы машинного обучения, в частности метод Adaptive Boosting (на основе метода градиентного спуска), методы обучения нейронных сетей. **Результаты:** были использованы распределенные по областям Украины данные Центра общественного здоровья Минздрава Украины заболеваемостью COVID-19, количество лабораторно обследованных лиц, количество проведенных лабораторных исследований методами ПЦР и ИФА, количество проведенных лабораторных исследований IgA, IgM, IgG; в модели использованы данные с марта 2020 по декабрь 2020, при моделировании не учтены данные с временно оккупированных территорий Украины; для кластерного анализа построена нейронная сеть из 60 входных нейронов, 100 скрытых нейронов с активационной функцией Ферми и 4 выходными нейронами; для программной реализации модели использован язык программирования Python. **Выводы:** проведен анализ методов построения нейронных сетей; анализ методов обучения нейронных сетей, в том числе применения метода градиентного спуска для метода Adaptive Boosting; все теоретические сведения, описанные в данной работе, были использованы для реализации программного продукта обработки данных тестирования на COVID-19 в Украине; было проведено разбиение областей Украины на зоны заражения вирусом COVID-19 и представлена карта этого разбиения.

**Ключевые слова:** кластерный анализ; нейронная сеть; машинное обучение; эпидемический процесс; COVID-19.

## IDENTIFICATION OF AREAS OF CORONAVIRUS COVID-19 INCIDENCE SPREADING BASED ON CLUSTER ANALYSIS METHOD

**Subject:** the use of the mathematical apparatus of neural networks for the scientific substantiation of anti-epidemic measures in order to reduce the incidence of diseases when making effective management decisions. **Purpose:** to apply cluster analysis, based on a neural network, to solve the problem of identifying areas of incidence. **Tasks:** to analyze methods of data analysis to solve the clustering problem; to develop a neural network method for clustering the territory of Ukraine according to the nature of the epidemic process COVID-19; on the basis of the developed method, to implement a data analysis software product to identify the areas of incidence of the disease using the example of the coronavirus COVID-19. **Methods:** models and methods of data analysis, models and methods of systems theory (based on the information approach), machine learning methods, in particular the Adaptive Boosting method (based on the gradient descent method), methods for training neural networks. **Results:** we used the data of the Center for Public Health of the Ministry of Health of Ukraine distributed over the regions of Ukraine on the incidence of COVID-19, the number of laboratory examined persons, the number of laboratory tests performed by PCR and ELISA methods, the number of laboratory tests of IgA, IgM, IgG; the model used data from March 2020 to December 2020, the modeling did not take into account data from the temporarily occupied territories of Ukraine; for cluster analysis, a neural network of 60 input neurons, 100 hidden neurons with an activation Fermi function and 4 output neurons was built; for the software implementation of the model, the programming language Python was used. **Conclusions:** analysis of methods for constructing neural networks; analysis of training methods for neural networks, including the use of the gradient descent method for the Adaptive Boosting method; all theoretical information described in this work was used to implement a software product for processing test data for COVID-19 in Ukraine; the division of the regions of Ukraine into zones of infection with the COVID-19 virus was carried out and a map of this division was presented.

**Keywords:** cluster analysis; neural network; machine learning; epidemic process; COVID-19.

### *Бібліографічні описи / Bibliographic descriptions*

Базілевич К. О., Меньяйлов Є. С., Чумаченко Д. І. Виділення зон розповсюдження захворюваності на коронавірус COVID-19 на основі методів кластерного аналізу. *Сучасний стан наукових досліджень та технологій в промисловості*. 2021. № 1 (15). С. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2021.15.005>

Bazilevych, K., Menailov, Ie., Chumachenko, D. (2021), "Identification of areas of coronavirus COVID-19 incidence spreading based on cluster analysis method", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (15), P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2021.15.005>